On the Variability of the Sample Covariance Matrix Under Complex Elliptical Distributions

Elias Raninen , *Student Member, IEEE*, Esa Ollila , *Senior Member, IEEE*, and David E. Tyler

Abstract—We derive the form of the variance-covariance matrix for any affine equivariant matrix-valued statistics when sampling from complex elliptical distributions. We then use this result to derive the variance-covariance matrix of the sample covariance matrix (SCM) as well as its theoretical mean squared error (MSE) when finite fourth-order moments exist. Finally, illustrative examples of the formulas are presented.

Index Terms—Sample covariance matrix, sample variation, mean squared error, complex Gaussian distribution, complex elliptically symmetric distribution.

I. INTRODUCTION

SUPPOSE we observe independent and identically distributed (i.i.d.) complex-valued p-variate random vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n \subset \mathbb{C}^p$ with mean $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}_i]$ and positive definite covariance matrix $\boldsymbol{\Sigma} = \mathbb{E}[(\mathbf{x}_i - \boldsymbol{\mu})(\mathbf{x}_i - \boldsymbol{\mu})^{\mathsf{H}}]$. The (unbiased) estimators of $\boldsymbol{\Sigma}$ and $\boldsymbol{\mu}$ are the sample covariance matrix (SCM) and the sample mean defined by

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^{n} (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathsf{H}} \text{ and } \bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i. \quad (1)$$

The SCM is an integral part of many statistical signal processing methods such as adaptive filtering (Wiener and Kalman filters), spectral estimation and array processing (MUSIC algorithm, Capon beamformer) [1], [2], and adaptive radar detectors [3]–[5].

In signal processing applications, a typical assumption would be to assume that the data $\mathbf{x}_1, \ldots, \mathbf{x}_n$ follow a (circular) complex multivariate normal (MVN) distribution [6], denoted by $\mathbb{CN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. However, a more general assumption would be to assume a Complex Elliptically Symmetric (CES) [7], [8] distribution, which is a family of distributions including the MVN distribution as well as heavier-tailed distributions such as the t-, K-, and the inverse Gaussian distribution that are commonly used in radar and array signal processing applications as special cases [8]–[11].

In the paper, we study the complex-valued (unbiased) SCM for which we derive the variance-covariance matrix as well

as the theoretical mean squared error (MSE) when sampling from CES distributions. We also provide a general expression for the variance-covariance matrix of any affine equivariant matrix-valued statistic (of which the SCM is a particular case). The results regarding the SCM extend the results in [12] to the complex-valued case, where the variance-covariance matrix and MSE of the SCM were derived for real-valued elliptical distributions.

The structure of the paper is as follows. Section II introduces CES distributions. In Section III, we derive the variance-covariance matrix of any affine equivariant matrix-valued statistic when sampling from a CES distribution. In Section IV, we derive the variance-covariance matrix of the SCM and provide an application in shrinkage estimation. Section V concludes. All proofs are kept in the appendix.

Notation: We let $\mathbf{I}$, $\mathbf{1}$, and $\mathbf{e}_i$ denote the identity matrix, a vector of ones, and a vector whose ith coordinate is one and other coordinates are zero, respectively. The notations $(\cdot)^*$, $(\cdot)^{\top}$, and $(\cdot)^{\mathsf{H}}$, denote the complex conjugate, the transpose, and the conjugate transpose, respectively. The notations $\mathcal{H}^p$, $\mathcal{H}^p_+$, and $\mathcal{H}^p_{++}$ denote the sets of Hermitian, Hermitian positive semidefinite, and Hermitian positive definite $p \times p$ dimensional matrices, respectively. We use the shorthand notation $\mathrm{var}(\mathbf{A}) = \mathrm{var}(\mathrm{vec}(\mathbf{A}))$ and $\mathrm{pvar}(\mathbf{A}) = \mathrm{pvar}(\mathrm{vec}(\mathbf{A}))$ (see Section III for the definition of pvar), where $\mathrm{vec}(\mathbf{A}) = (\mathbf{a}_1^{\top} \cdots \mathbf{a}_p^{\top})^{\top}$ is a vectorization of $\mathbf{A} = (\mathbf{a}_1 \cdots \mathbf{a}_p)$. When there is a possibility for confusion, we denote by $\mathrm{cov}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}(\cdot, \cdot)$ or $\mathbb{E}_{\boldsymbol{\mu}, \boldsymbol{\Sigma}}[\cdot]$ the covariance and expectation of a sample from an elliptical distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. The $p^2 \times p^2$ commutation matrix [13] is defined by $\mathbf{K}_{p,p} = \sum_{i,j} \mathbf{e}_i \mathbf{e}_j^{\top} \otimes \mathbf{e}_j \mathbf{e}_i^{\top}$, where $\otimes$ is the Kronecker product. The notation $\overset{d}{=}$ reads "has the same distribution as", $\mathcal{U}(\mathbb{C}\mathcal{S}^p)$ denotes the uniform distribution on the complex unit sphere $\mathbb{C}\mathcal{S}^p = \{\mathbf{u} \in \mathbb{C}^p : \|\mathbf{u}\| = 1\}$ and $\mathbb{R}_{\geq 0} = \{a \in \mathbb{R} : a \geq 0\}$.

II. COMPLEX ELLIPTICALLY SYMMETRIC DISTRIBUTIONS

A random vector $\mathbf{x} \in \mathbb{C}^p$ is said to have a circular CES distribution if and only if it admits the *stochastic representation*

$$\mathbf{x} \overset{d}{=} \boldsymbol{\mu} + r \boldsymbol{\Sigma}^{1/2} \mathbf{u}, \quad (2)$$

where $\boldsymbol{\mu} = \mathbb{E}[\mathbf{x}]$ is the mean vector, $\boldsymbol{\Sigma}^{1/2} \in \mathcal{H}^p_{++}$ is the unique Hermitian positive definite square-root of $\boldsymbol{\Sigma}$, $\mathbf{u} \sim \mathcal{U}(\mathbb{C}\mathcal{S}^p)$, and $r > 0$ is a positive random variable called the *modular variate*. Furthermore r and $\mathbf{u}$ are independent. If the cumulative distribution function of $\mathbf{x}$ is absolutely continuous, the probability density function exists and is up to a constant of the form

$$|\boldsymbol{\Sigma}|^{-1} g((\mathbf{x} - \boldsymbol{\mu})^{\mathsf{H}} \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})), \quad (3)$$

where $g : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is the *density generator*. We denote this case by $\mathbf{x} \sim \mathbb{C}\mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. We assume that $\mathbf{x}$ has finite fourth-order moments, and thus we can assume without any loss of generality that $\boldsymbol{\Sigma}$ is equal to the covariance matrix $\mathrm{var}(\mathbf{x})$ [8] (implying $\mathbb{E}[r^2] = p$). We refer the reader to [8] for a comprehensive account on CES distributions.

The elliptical kurtosis of a CES distribution is defined as

$$\kappa = \frac{\mathbb{E}[r^4]}{p(p+1)} - 1. \qquad (4)$$

Elliptical kurtosis shares properties similar to the kurtosis of a circular complex random variable. Specifically, if $\mathbf{x} \sim \mathbb{C}\mathcal{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then $\kappa = 0$. This follows by noticing that $2 \cdot r^2 \sim \chi^2_{2p}$, and hence $\mathbb{E}[r^4] = p(p+1)$ and consequently $\kappa = 0$ in the Gaussian case. The kurtosis of a complex circularly symmetric random variable $x \in \mathbb{C}$ is defined as

$$\mathrm{kurt}(x) = \frac{\mathbb{E}[|x - \mu|^4]}{(\mathbb{E}[|x - \mu|^2])^2} - 2, \qquad (5)$$

where $\mu = \mathbb{E}[x]$. Similar to the real-valued case, κ has a simple relationship with the (excess) kurtosis [14, Lemma 3]: $\kappa = \frac{1}{2} \cdot \mathrm{kurt}(x_i)$, for any $i \in \{1, \ldots, p\}$. We note that the lower bound for the elliptical kurtosis is $\kappa \geq -1/(p+1)$ [8].

Lastly, we define the *scale* and *sphericity* parameters

$$\eta = \frac{\mathrm{tr}(\boldsymbol{\Sigma})}{p} \quad \text{and} \quad \gamma = p \frac{\mathrm{tr}(\boldsymbol{\Sigma}^2)}{\mathrm{tr}(\boldsymbol{\Sigma})^2}. \qquad (6)$$

The scale is equal to the mean of the eigenvalues. The sphericity measures how close the covariance matrix is to a scaled identity matrix. The sphericity parameter gets the value 1 for the scaled identity matrix and p for a rank one matrix.

III. RADIAL DISTRIBUTIONS AND COVARIANCE MATRIX ESTIMATES

In this section, we derive the variance-covariance matrix of any affine equivariant matrix-valued statistic. We begin with some definitions.

The covariance and pseudo-covariance [15] of complex random vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ are defined as

$$\mathrm{cov}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{E}\left[(\mathbf{x}_1 - \mathbb{E}[\mathbf{x}_1])(\mathbf{x}_2 - \mathbb{E}[\mathbf{x}_2])^{\mathsf{H}}\right] \text{ and}$$

$$\mathrm{pcov}(\mathbf{x}_1, \mathbf{x}_2) = \mathbb{E}\left[(\mathbf{x}_1 - \mathbb{E}[\mathbf{x}_1])(\mathbf{x}_2 - \mathbb{E}[\mathbf{x}_2])^{\top}\right],$$

and together they provide a complete second-order description of associations between $\mathbf{x}_1$ and $\mathbf{x}_2$. Then $\mathrm{var}(\mathbf{x}) = \mathrm{cov}(\mathbf{x}, \mathbf{x})$ and $\mathrm{pvar}(\mathbf{x}) = \mathrm{pcov}(\mathbf{x}, \mathbf{x})$ are called the covariance matrix and the *pseudo-covariance matrix* [15] of $\mathbf{x}$.

A random Hermitian ($\mathbf{A}^{\mathsf{H}} = \mathbf{A}$) matrix $\mathbf{A} \in \mathcal{H}^p$ is said to have a *radial distribution* if $\mathbf{A} \stackrel{d}{=} \mathbf{Q}\mathbf{A}\mathbf{Q}^{\mathsf{H}}$ for all unitary matrices $\mathbf{Q}$ (so $\mathbf{Q}^{\mathsf{H}}\mathbf{Q} = \mathbf{I}$). The following result extends the result of [16] to the complex-valued case.

Theorem 1: Let a random matrix $\mathbf{A} = (a_{ij}) \in \mathcal{H}^p$ have a radial distribution with finite second-order moments. Then, there exist real-valued constants σ, τ_1 and τ_2 with $\tau_1 \geq 0$ and $\tau_2 \geq -\tau_1/p$ such that $\mathbb{E}[\mathbf{A}] = \sigma\mathbf{I}$ with $\sigma = \mathbb{E}[a_{ii}]$ and

$$\mathrm{var}(\mathbf{A}) = \tau_1\mathbf{I} + \tau_2 \, \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^{\top}, \qquad (7)$$

$$\mathrm{pvar}(\mathbf{A}) = \tau_1\mathbf{K}_{p,p} + \tau_2 \, \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^{\top}, \qquad (8)$$

where $\tau_1 = \mathrm{var}(a_{ij}) = \mathrm{pcov}(a_{ij}, a_{ji})$ and $\tau_2 = \mathrm{cov}(a_{ii}, a_{jj}) = \mathrm{pcov}(a_{ii}, a_{jj})$ for all $1 \leq i \neq j \leq p$.

A statistic $\hat{\boldsymbol{\Sigma}} = \hat{\boldsymbol{\Sigma}}(\mathbf{X}) \in \mathcal{H}^p$ based on an $n \times p$ data matrix $\mathbf{X} = \begin{pmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_n \end{pmatrix}^{\top}$ of $n \geq 1$ observations on p complex-valued variables is said to be *affine equivariant* if

$$\hat{\boldsymbol{\Sigma}}(\mathbf{X}\mathbf{A}^{\top} + \mathbf{1}\mathbf{a}^{\top}) = \mathbf{A}\hat{\boldsymbol{\Sigma}}(\mathbf{X})\mathbf{A}^{\mathsf{H}} \qquad (9)$$

holds for all $\mathbf{A} \in \mathbb{C}^{p \times p}$ and $\mathbf{a} \in \mathbb{C}^p$. Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n \subset \mathbb{C}^p$ is a random sample from a CES distribution $\mathbb{C}\mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ and that $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij}) \in \mathcal{H}^p$ is an affine equivariant statistic. Then $\hat{\boldsymbol{\Sigma}}$ has a stochastic decomposition

$$\hat{\boldsymbol{\Sigma}}(\mathbf{X}) \stackrel{d}{=} \boldsymbol{\Sigma}^{1/2} \cdot \hat{\boldsymbol{\Sigma}}(\mathbf{Z}) \cdot \boldsymbol{\Sigma}^{1/2}, \qquad (10)$$

where $\hat{\boldsymbol{\Sigma}}(\mathbf{Z})$ denotes the value of $\hat{\boldsymbol{\Sigma}}$ based on a random sample $\mathbf{z}_1, \ldots, \mathbf{z}_n \subset \mathbb{C}^p$ from a spherical distribution $\mathbb{C}\mathcal{E}_p(\mathbf{0}, \mathbf{I}, g)$. This follows by writing $\mathbf{X} \stackrel{d}{=} \mathbf{Z}(\boldsymbol{\Sigma}^{1/2})^{\top} + \mathbf{1}\boldsymbol{\mu}^{\top}$ using (2) (where $\mathbf{z}_i = r_i\mathbf{u}_i$) and then applying (9). Affine equivariance together with the fact that $\mathbf{z}_i \stackrel{d}{=} \mathbf{Q}\mathbf{z}_i$ for all unitary matrices $\mathbf{Q}$ indicate that $\hat{\boldsymbol{\Sigma}}(\mathbf{Z})$ has a radial distribution. This leads to Theorem 2 stated below.

Theorem 2: Let $\hat{\boldsymbol{\Sigma}} = (\hat{\sigma}_{ij}) \in \mathcal{H}^p$ be an affine equivariant statistic with finite second-order moments, and based on a random sample $\mathbf{x}_1, \ldots, \mathbf{x}_n \subset \mathbb{C}^p$ from a CES distribution $\mathbb{C}\mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$. Then $\mathbb{E}[\hat{\boldsymbol{\Sigma}}] = \sigma\boldsymbol{\Sigma}$ with $\sigma = \mathbb{E}_{\mathbf{0}, \mathbf{I}}[\hat{\sigma}_{11}]$, and

$$\mathrm{var}(\hat{\boldsymbol{\Sigma}}) = \tau_1(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}) + \tau_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathsf{H}}, \qquad (11)$$

$$\mathrm{pvar}(\hat{\boldsymbol{\Sigma}}) = \tau_1(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma})\mathbf{K}_{p,p} + \tau_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\top}, \qquad (12)$$

where $\tau_1 = \mathrm{var}_{\mathbf{0}, \mathbf{I}}(\hat{\sigma}_{12})$ and $\tau_2 = \mathrm{cov}_{\mathbf{0}, \mathbf{I}}(\hat{\sigma}_{11}, \hat{\sigma}_{22}) \geq -\tau_1/p$.

There are many statistics for which this theorem applies. Naturally, a prominent example is the SCM, which we examine in detail in the next section. Other examples are the complex M-estimates of scatter discussed in [8] or the weighted sample covariance matrices $\mathbf{R} = \frac{1}{n} \sum_{i=1}^{n} u(d_i)(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathsf{H}}$, where $d_i = (\mathbf{x}_i - \bar{\mathbf{x}})^{\mathsf{H}}\mathbf{S}^{-1}(\mathbf{x}_i - \bar{\mathbf{x}})$ and $u : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$. In the special case, when $u(s) = s$, we obtain the fourth moment matrix used in FOBI [17] for blind source separation and in Invariant Coordinate Selection (ICS) [18].

IV. VARIANCE-COVARIANCE OF THE SCM AND AN EXAMPLE IN SHRINKAGE ESTIMATION

We now use Theorem 2 to derive the covariance matrix and the pseudo-covariance matrix as well as the MSE of the SCM when sampling from a CES distribution. This result extends [12, Theorem 2 and Lemma 1] to the complex case.

Theorem 3: Let the SCM $\mathbf{S} = (s_{ij})$ be computed on an i.i.d. random sample $\mathbf{x}_1, \ldots, \mathbf{x}_n \subset \mathbb{C}^p$ from a CES distribution $\mathbb{C}\mathcal{E}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g)$ with finite fourth-order moments and covariance matrix $\boldsymbol{\Sigma} = \mathrm{var}(\mathbf{x}_i)$. Then, the covariance matrix and pseudo-covariance matrix of $\mathbf{S}$ are as stated in (11) and (12) with

$$\tau_1 = \mathrm{var}_{\mathbf{0}, \mathbf{I}}(s_{12}) = \frac{1}{n-1} + \frac{\kappa}{n},$$

$$\tau_2 = \mathrm{cov}_{\mathbf{0}, \mathbf{I}}(s_{11}, s_{22}) = \frac{\kappa}{n},$$

where κ is the elliptical kurtosis in (4). The MSE is given by

$$\mathrm{MSE}(\mathbf{S}) = \mathbb{E}[\|\mathbf{S} - \boldsymbol{\Sigma}\|_{\mathrm{F}}^2] = \left(\frac{1}{n-1} + \frac{\kappa}{n}\right)\mathrm{tr}(\boldsymbol{\Sigma})^2 + \frac{\kappa}{n}\mathrm{tr}(\boldsymbol{\Sigma}^2),$$

and the normalized MSE is

$$\mathrm{NMSE}(\mathbf{S}) = \frac{\mathrm{MSE}(\mathbf{S})}{\|\boldsymbol{\Sigma}\|_{\mathrm{F}}^2} = \frac{p}{\gamma}\left(\frac{1}{n-1} + \frac{\kappa}{n}\right) + \frac{\kappa}{n}, \qquad (13)$$
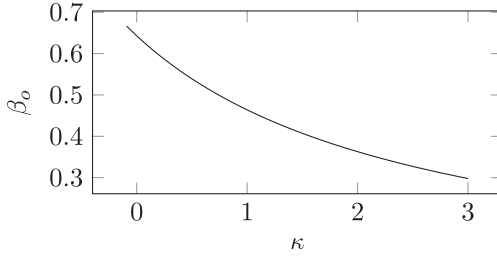
Fig. 1.   The value of β_o as a function of κ, when $n = p = 10$ and $\gamma = 2$.

where γ is the sphericity parameter defined in (6).

The finite sample performance of the SCM can often be improved by using shrinkage covariance matrix estimators, which is an active area of research, see e.g., [12], [19]–[24]. Consider the simple shrinkage covariance matrix estimation problem,

$$\beta_o = \arg\min_{\beta \in \mathbb{R}} \mathbb{E}[\|\beta \mathbf{S} - \mathbf{\Sigma}\|_{\mathrm{F}}^2].$$

Since the problem is convex, we can find β_o as solution of $\frac{\partial}{\partial \beta} \mathbb{E}[\|\beta \mathbf{S} - \mathbf{\Sigma}\|_{\mathrm{F}}^2] = 2\beta \, \mathbb{E}[\mathrm{tr}(\mathbf{S}^2)] - 2\|\mathbf{\Sigma}\|_{\mathrm{F}}^2 = 0$ which yields

$$\beta_o = \frac{\|\mathbf{\Sigma}\|_{\mathrm{F}}^2}{\mathrm{MSE}(\mathbf{S}) + \|\mathbf{\Sigma}\|_{\mathrm{F}}^2} = \frac{1}{\mathrm{NMSE}(\mathbf{S}) + 1}, \quad (14)$$

where we used $\mathrm{MSE}(\mathbf{S}) = \mathbb{E}[\|\mathbf{S} - \mathbf{\Sigma}\|_{\mathrm{F}}^2] = \mathbb{E}[\mathrm{tr}(\mathbf{S}^2)] - \|\mathbf{\Sigma}\|_{\mathrm{F}}^2$. From (14), it follows that the optimal scaling term β_o is always smaller than 1 since $\mathrm{MSE}(\mathbf{S}) > 0$. Furthermore, β_o is a function of γ and κ via (13). In the Gaussian case ($\kappa = 0$), we obtain $\beta_o = (n-1)/(n-1+p/\gamma)$. Fig. 1 illustrates the effect of κ on β_o when $\gamma = 2$, and $n = p = 10$. Next we show that the oracle estimator $\mathbf{S}_o = \beta_o \mathbf{S}$ is uniformly more efficient than $\mathbf{S}$, i.e., $\mathrm{MSE}(\mathbf{S}_0) < \mathrm{MSE}(\mathbf{S})$ for any $\mathbf{\Sigma} \in \mathcal{H}_{++}^p$. First write

$$\mathbb{E}\left[\|\beta_o \mathbf{S} - \mathbf{\Sigma}\|_{\mathrm{F}}^2\right] = \beta_o^2 \mathrm{MSE}(\mathbf{S}) + (1 - \beta_o)^2 \|\mathbf{\Sigma}\|_{\mathrm{F}}^2. \quad (15)$$

Then from (14) we notice that $1 - \beta_o = \beta_o \mathrm{NMSE}(\mathbf{S})$. Subsituting this into (15) we get

$$\mathrm{MSE}(\mathbf{S}_o) = \beta_o^2 \mathrm{MSE}(\mathbf{S}) + \beta_o^2 \mathrm{NMSE}(\mathbf{S})^2 \|\mathbf{\Sigma}\|_{\mathrm{F}}^2$$

$$= \beta_o^2 \mathrm{MSE}(\mathbf{S})\left(1 + \mathrm{NMSE}(\mathbf{S})\right) = \beta_o \mathrm{MSE}(\mathbf{S}),$$

where the last identity follows from fact that $1/\beta_o = 1 + \mathrm{NMSE}(\mathbf{S})$ due to (14). Since $\beta_o < 1$ for all $\mathbf{\Sigma} \in \mathcal{H}_{++}^p$, it follows that $\mathbf{S}_o$ is more efficient than $\mathbf{S}$. Efficiency in the case when γ and κ, and hence β_o need to be estimated, remains (to the best of our knowledge) an open problem. However, for certain related shrinkage estimators the shrinkage intensity can be consistently estimated, e.g., [19], [20].

In the univariate case ($p = 1$), $\mathbf{\Sigma}$ is equal to the variance $\sigma^2 = \mathrm{var}(x) > 0$ of the random variable $x \in \mathbb{C}$ and the SCM reduces to the sample variance defined by $s^2 = \frac{1}{n-1} \sum_{i=1}^n |x_i - \bar{x}|^2$. In this case, $\gamma = 1$, and β_o in (14) is

$$\beta_0 = \frac{n(n-1)}{\mathrm{kurt}(x)(n-1) + n^2}.$$

A similar result was noticed in [25] for the real-valued case. If the data is from a complex normal distribution $\mathbb{CN}(\mu, \sigma^2)$, then $\mathrm{kurt}(x) = 0$, and $\beta_o = (n-1)/n$, and hence $s_o^2 = \beta_o s^2 = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}|^2$, which equals the Maximum Likelihood Estimate (MLE) of σ^2. In the real case, the optimal scaling constant is $\beta_o = (n-1)/(n+1)$ for Gaussian samples [26]. Note that when the kurtosis is large and positive and n is small, then β_o can be substantially less than one and the gain of using s_o can be significant.

V. CONCLUSION

We derived the form of the variance-covariance matrix for any affine equivariant matrix-valued statistics under sampling from CES distributions. We used this result to derive the variance-covariance matrix and the MSE of the SCM when finite fourth-order moments exist. An illustrative example in the context of shrinkage covariance matrix estimation was presented.

APPENDIX

A. Proof of Theorem 1

The proof follows the same lines as the proof in [16] for the real-valued case. Since $\mathbf{A}$ has a radial distribution $\mathbb{E}[\mathbf{A}] = \mathbb{E}[a_{ii}]\mathbf{I} = \sigma \mathbf{I}$ is obvious. For any unitary matrix $\mathbf{Q} = (\mathbf{q}_1 \cdots \mathbf{q}_p)$, we have

$$\mathrm{var}(\mathbf{A}) = \mathrm{var}(\mathrm{vec}(\mathbf{A})) = \mathrm{var}(\mathrm{vec}(\mathbf{Q}\mathbf{A}\mathbf{Q}^{\mathsf{H}}))$$

$$= \mathrm{var}((\mathbf{Q}^* \otimes \mathbf{Q})\mathrm{vec}(\mathbf{A}))$$

$$= (\mathbf{Q}^* \otimes \mathbf{Q}) \, \mathrm{var}(\mathrm{vec}(\mathbf{A}))(\mathbf{Q}^\top \otimes \mathbf{Q}^{\mathsf{H}}).$$

Let $\{\mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_k\mathbf{e}_l^\top\}$ be a basis for the set of $p^2 \times p^2$ matrices. Then

$$\mathrm{var}(\mathbf{A}) = \sum_{i,j,k,l} \tau_{ijkl}\mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_k\mathbf{e}_l^\top = \sum_{i,j,k,l} \tau_{ijkl}\mathbf{q}_i^*\mathbf{q}_j^\top \otimes \mathbf{q}_k\mathbf{q}_l^{\mathsf{H}},$$

where $\tau_{ijkl} = \mathrm{var}(a_{ki}, a_{lj})$. By choosing $\mathbf{q}_m = \jmath\mathbf{e}_m$ (where $\jmath$ is the imaginary unit) and $\mathbf{q}_r = \mathbf{e}_r$ for some $m \neq r$, we must have $\tau_{ijkl} = 0$ unless $i = j = k = l$, $i = j \neq k = l$, or $i = k \neq j = l$. Denote $\tau_0 = \tau_{iiii} = \mathrm{var}(a_{ii})$, $\tau_1 = \tau_{iijj} = \mathrm{var}(a_{ji})$ and $\tau_2 = \tau_{ijij} = \mathrm{cov}(a_{ii}, a_{jj})$. Then

$$\mathrm{var}(\mathbf{A}) = \sum_{i,j} \tau_1 \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_j\mathbf{e}_j^\top + \sum_{i,j} \tau_2 \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_i\mathbf{e}_j^\top$$

$$+ (\tau_0 - \tau_1 - \tau_2) \sum_i \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top.$$

Note that $\sum_{i,j} \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_j\mathbf{e}_j^\top = \mathbf{I}$ and $\sum_{i,j} \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_i\mathbf{e}_j^\top = \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top$. Furthermore,

$$(\mathbf{Q}^* \otimes \mathbf{Q}) \sum_{i,j} \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_j\mathbf{e}_j^\top \left(\mathbf{Q}^\top \otimes \mathbf{Q}^{\mathsf{H}}\right) = \mathbf{I}$$

$$(\mathbf{Q}^* \otimes \mathbf{Q}) \sum_{i,j} \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_i\mathbf{e}_j^\top \left(\mathbf{Q}^\top \otimes \mathbf{Q}^{\mathsf{H}}\right) = \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top$$

$$(\mathbf{Q}^* \otimes \mathbf{Q}) \sum_i \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top \left(\mathbf{Q}^\top \otimes \mathbf{Q}^{\mathsf{H}}\right) \neq \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_i\mathbf{e}_j^\top.$$

From the last inequality, we must have $\tau_0 - \tau_1 - \tau_2 = 0$ and $\mathrm{var}(\mathbf{A}) = \tau_1 \mathbf{I} + \tau_2 \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top$ follows.

Regarding the pseudo-covariance, for any unitary $\mathbf{Q}$,

$$\mathrm{pvar}(\mathbf{A}) = (\mathbf{Q}^* \otimes \mathbf{Q}) \, \mathrm{pvar}(\mathrm{vec}(\mathbf{A}))(\mathbf{Q}^{\mathsf{H}} \otimes \mathbf{Q}^\top),$$

which implies

$$\mathrm{pvar}(\mathbf{A}) = \sum_{i,j,k,l} \tau'_{ijkl}\mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_k\mathbf{e}_l^\top = \sum_{i,j,k,l} \tau'_{ijkl}\mathbf{q}_i^*\mathbf{q}_j^{\mathsf{H}} \otimes \mathbf{q}_k\mathbf{q}_l^\top,$$

where $\tau'_{ijkl} = \mathrm{pcov}(a_{ki}, a_{lj})$. By choosing $\mathbf{q}_m = \jmath\mathbf{e}_m$ and $\mathbf{q}_r = \mathbf{e}_r$ for some $m \neq r$, we must have $\tau'_{ijkl} = 0$ except when $i = j = k = l$, $i = k \neq j = l$, or $i = l \neq j = k$. Let $\tau'_0 = \tau'_{iiii} = \mathrm{pvar}(a_{ii})$, $\tau'_1 = \tau'_{ijji} = \mathrm{pcov}(a_{ij}, a_{ji})$ and $\tau'_2 = \tau'_{ijij} = \mathrm{pcov}(a_{ii}, a_{jj})$. Then,

$$\mathrm{pvar}(\mathbf{A}) = \sum_{i,j} \tau'_1 \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_j\mathbf{e}_i^\top + \sum_{i,j} \tau'_2 \mathbf{e}_i\mathbf{e}_j^\top \otimes \mathbf{e}_i\mathbf{e}_j^\top$$

$$+ (\tau'_0 - \tau'_1 - \tau'_2) \sum_i \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top$$

$$= \tau'_1 \mathbf{K}_{p,p} + \tau'_2 \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top$$

by similar arguments as with $\mathrm{var}(\mathbf{A})$. Then note that, $\tau_1 = \mathrm{var}(a_{ji}) = \mathrm{pcov}(a_{ij}, a_{ji}) = \tau'_1 \geq 0$ and $\tau_2 = \mathrm{cov}(a_{ii}, a_{jj}) = \mathrm{pcov}(a_{ii}, a_{jj}) = \tau'_2$. Lastly, since $\mathrm{var}(\mathbf{A})$ is positive semidefinite and $\tau_1 \geq 0$, $|\mathrm{var}(\mathbf{A})| = |\tau_1 \mathbf{I} + \tau_2 \mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top| = (\tau_1 + \tau_2 p)\tau_1^{p^2-1} \geq 0$ implies $\tau_2 \geq -\tau_1/p$. $\square$

B. Proof of Theorem 2

Since $\hat{\boldsymbol{\Sigma}}(\mathbf{X})$ is affine equivariant, from (10) we have

$$\mathrm{var}(\hat{\boldsymbol{\Sigma}}(\mathbf{X})) = \mathrm{var}(\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{\Sigma}}(\mathbf{Z})\boldsymbol{\Sigma}^{1/2})$$

$$= ((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\mathrm{var}(\hat{\boldsymbol{\Sigma}}(\mathbf{Z}))((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2}).$$

From Theorem 1, $\mathrm{var}(\hat{\boldsymbol{\Sigma}}(\mathbf{Z}))$ is of the form (7). Since

$$((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\mathbf{I}((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2}) = (\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}) \text{ and }$$

$$((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})$$

$$= \mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathsf{H}},$$

we obtain (11). Similarly,

$$\mathrm{pvar}(\hat{\boldsymbol{\Sigma}}(\mathbf{X})) = \mathrm{pvar}(\boldsymbol{\Sigma}^{1/2}\hat{\boldsymbol{\Sigma}}(\mathbf{Z})\boldsymbol{\Sigma}^{1/2})$$

$$= ((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\,\mathrm{pvar}(\hat{\boldsymbol{\Sigma}}(\mathbf{Z}))(\boldsymbol{\Sigma}^{1/2} \otimes (\boldsymbol{\Sigma}^{1/2})^*),$$

where $\mathrm{pvar}(\hat{\boldsymbol{\Sigma}}(\mathbf{Z}))$ is of the form (8). Since

$$((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\mathbf{K}_{p,p}(\boldsymbol{\Sigma}^{1/2} \otimes (\boldsymbol{\Sigma}^{1/2})^*) = (\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma})\mathbf{K}_{p,p},$$

$$((\boldsymbol{\Sigma}^{1/2})^* \otimes \boldsymbol{\Sigma}^{1/2})\mathrm{vec}(\mathbf{I})\mathrm{vec}(\mathbf{I})^\top(\boldsymbol{\Sigma}^{1/2} \otimes (\boldsymbol{\Sigma}^{1/2})^*)$$

$$= \mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^\top,$$

we obtain (12). $\square$

C. Proof of Theorem 3

The proof is similar to the proof of [12, Theorem 2] that was derived for the real-valued case. Write the SCM as $\mathbf{S} = (s_{ij}) = (n-1)^{-1}\mathbf{X}^\top\mathbf{H}\mathbf{X}^*$, where $\mathbf{H} = \mathbf{I} - \frac{1}{n}\mathbf{1}\mathbf{1}^\top$ is the centering matrix. Write $\mathbf{a} = \mathbf{X}\mathbf{e}_q$ and $\mathbf{b} = \mathbf{X}\mathbf{e}_r$ for $q \neq r$. Then note that $s_{qr} = (n-1)^{-1}\mathbf{a}^\top\mathbf{H}\mathbf{b}^*$. Hence,

$$\tau_1 = \mathrm{var}(s_{qr}) = \mathrm{var}((n-1)^{-1}\mathbf{a}^\top\mathbf{H}\mathbf{b}^*)$$

$$= (n-1)^{-2}\,\mathrm{var}(\mathbf{a}^\top\mathbf{H}\mathbf{b}^*). \tag{16}$$

Then note that

$$\mathrm{var}(\mathbf{a}^\top\mathbf{H}\mathbf{b}^*) = \mathrm{var}(\mathrm{tr}(\mathbf{H}\mathbf{b}^*\mathbf{a}^\top)) \tag{17}$$

$$= \mathrm{var}(\mathrm{vec}(\mathbf{H})^\top \mathrm{vec}(\mathbf{b}^*\mathbf{a}^\top))$$

$$= \mathrm{vec}(\mathbf{H})^\top \mathrm{var}(\mathrm{vec}(\mathbf{b}^*\mathbf{a}^\top))\mathrm{vec}(\mathbf{H}). \tag{18}$$

Since for $\mathbf{x}_i \sim \mathbb{C}\mathcal{E}_p(\mathbf{0}, \mathbf{I}, g)$ we have $\mathbf{x}_i \stackrel{d}{=} r_i\mathbf{u}_i$, where $r_i \stackrel{d}{=} \|\mathbf{x}_i\|$ is independent of $\mathbf{u}_i = (u_{i1}, \ldots, u_{ip})^\top \sim \mathcal{U}(\mathbb{C}\mathcal{S}^p)$, we can write $\mathbf{a} = \mathbf{X}\mathbf{e}_q = (r_1 u_{1q}, r_2 u_{2q}, \ldots, r_n u_{nq})^\top$, and similarly for $\mathbf{b}$. The klth element of the ijth block (i.e., the $ijkl$th element) of the $n^2 \times n^2$ matrix $\mathrm{var}(\mathrm{vec}(\mathbf{b}^*\mathbf{a}^\top))$ is

$$\mathrm{cov}\,(b_k^* a_i, b_l^* a_j) = \mathbb{E}\left[r_k r_i r_l r_j u_{kr}^* u_{iq} u_{lr} u_{jq}^*\right],$$

where we used that $\mathbb{E}[\mathbf{b}^*\mathbf{a}^\top] = \mathbf{0}$. Then note that

$$\mathbb{E}\left[|u_{iq}|^2 |u_{ir}|^2\right] = \frac{1}{p(p+1)},$$

$$\mathbb{E}\left[|u_{iq}|^2\right] = \frac{1}{p}, \text{ and } \mathbb{E}\left[|u_{iq}|^4\right] = \frac{2}{p(p+1)},$$

while all other moments up to fourth-order vanish. This and the fact that $\mathbb{E}[r_i^4] = (1+\kappa)p(p+1)$ due to (4), implies that the only non-zero elements of $\mathrm{var}(\mathrm{vec}(\mathbf{b}^*\mathbf{a}^\top))$ are

$$\mathbb{E}[r_i^4]\,\mathbb{E}[|u_{ir}|^2|u_{iq}|^2] = 1+\kappa \quad \text{for } i = j = k = l, \text{ and}$$

$$\mathbb{E}[r_i^2]\,\mathbb{E}[r_k^2]\,\mathbb{E}[u_{ir}^2]\,\mathbb{E}[u_{kq}^2] = 1 \qquad \text{for } i = j \neq k = l,$$

and hence

$$\mathrm{var}\left(\mathrm{vec}\left(\mathbf{b}^*\mathbf{a}^\top\right)\right) = \mathbf{I} + \kappa \sum_{i=1}^n \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top. \tag{19}$$

This together with (16) and (18) yields

$$\tau_1 = \frac{1}{(n-1)^2}\mathrm{vec}(\mathbf{H})^\top\left(\mathbf{I} + \kappa \sum_{i=1}^n \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top\right)\mathrm{vec}(\mathbf{H})$$

$$= \frac{1}{n-1} + \frac{\kappa}{n},$$

where we used $\mathrm{vec}(\mathbf{H})^\top\mathrm{vec}(\mathbf{H}) = n-1$ and

$$\sum_{i=1}^n \mathrm{vec}(\mathbf{H})^\top(\mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top)\mathrm{vec}(\mathbf{H}) = \sum_{i=1}^n h_{ii}^2 = \frac{(n-1)^2}{n}.$$

Next, we find the expression for $\tau_2 = \mathrm{cov}(s_{qq}, s_{rr}) = (n-1)^{-2}\,\mathrm{cov}(\mathbf{a}^\top\mathbf{H}\mathbf{a}^*, \mathbf{b}^\top\mathbf{H}\mathbf{b}^*)$. Since $\mathbb{E}[s_{qq}] = \mathbb{E}[(n-1)^{-1}\mathbf{a}^\top\mathbf{H}\mathbf{a}^*] = 1$ for any q,

$$\tau_2 = (n-1)^{-2}\,\mathbb{E}[\mathbf{a}^\top\mathbf{H}\mathbf{a}^*\mathbf{b}^\top\mathbf{H}\mathbf{b}^*] - 1$$

$$= (n-1)^{-2}\mathrm{tr}\left(\mathbb{E}\left[\mathbf{H}(\mathbf{b}^*\mathbf{a}^\top)^{\mathsf{H}}\mathbf{H}(\mathbf{b}^*\mathbf{a}^\top)\right]\right) - 1$$

$$= (n-1)^{-2}\mathrm{tr}\left((\mathbf{H}\otimes\mathbf{H})\,\mathbb{E}\left[\mathrm{vec}\left(\mathbf{b}^*\mathbf{a}^\top\right)\mathrm{vec}\left(\mathbf{b}^*\mathbf{a}^\top\right)^{\mathsf{H}}\right]\right) - 1.$$

The expression in the expectation is equal to (19), and so

$$\tau_2 = \frac{1}{(n-1)^2}\mathrm{tr}\left((\mathbf{H}\otimes\mathbf{H})\left(\mathbf{I} + \kappa\sum_{i=1}^n \mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top\right)\right) - 1$$

$$= \frac{\kappa}{n},$$

where we used that $\mathrm{tr}(\mathbf{H}\otimes\mathbf{H}) = \mathrm{tr}(\mathbf{H})^2 = (n-1)^2$ and

$$\sum_{i=1}^n \mathrm{tr}\left((\mathbf{H}\otimes\mathbf{H})\left(\mathbf{e}_i\mathbf{e}_i^\top \otimes \mathbf{e}_i\mathbf{e}_i^\top\right)\right) = \sum_{i=1}^n h_{ii}^2.$$

This completes the proof for τ_1 and τ_2. By Theorem 2, we have $\mathrm{var}(\mathbf{S}) = \tau_1(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}) + \tau_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathsf{H}}$, and hence

$$\mathrm{MSE}(\mathbf{S}) = \mathrm{tr}(\mathrm{var}(\mathbf{S}))$$

$$= \mathrm{tr}\left(\tau_1(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}) + \tau_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^{\mathsf{H}}\right)$$

$$= \tau_1\mathrm{tr}(\boldsymbol{\Sigma})^2 + \tau_2\mathrm{tr}(\boldsymbol{\Sigma}^2),$$

where the last identity follows from using $\mathrm{tr}(\boldsymbol{\Sigma}^* \otimes \boldsymbol{\Sigma}) = \mathrm{tr}(\boldsymbol{\Sigma})^2$. This gives the stated expression for the MSE. $\square$

REFERENCES

[1] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[2] D. G. Manolakis, V. K. Ingle, and S. M. Kogon, *Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing*. Norwood, MA, USA: Artech House, 2005.

[3] E. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, vol. AES-22, no. 2, pp. 115–127, Mar. 1986.

[4] F. Robey, D. Fuhrmann, E. Kelly, and R. Nitzberg, "A CFAR adaptive matched filter detector," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, no. 1, pp. 208–216, Jan. 1992.

[5] S. Kraut, L. L. Scharf, and L. T. McWhorter, "Adaptive subspace detectors," *IEEE Trans. Signal Process.*, vol. 49, no. 1, pp. 1–16, Jan. 2001.

[6] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (An introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[7] P. Krishnaiah and J. Lin, "Complex elliptically symmetric distributions," *Commun. Statist. - Theory Methods*, vol. 15, no. 12, pp. 3693–3718, Jan. 1986.

[8] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Complex elliptically symmetric distributions: Survey, new results and applications," *IEEE Trans. Signal Process.*, vol. 60, no. 11, pp. 5597–5625, Nov. 2012.

[9] E. Conte, M. Longo, and M. Lops, "Modelling and simulation of non-Rayleigh radar clutter," *IEE Proc. F-Radar Signal Process.*, vol. 138, no. 2, pp. 121–130, 1991.

[10] J. Billingsley, A. Farina, F. Gini, M. Greco, and L. Verrazzani, "Statistical analyses of measured radar ground clutter data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 2, pp. 579–593, Apr. 1999.

[11] E. Ollila, D. E. Tyler, V. Koivunen, and H. V. Poor, "Compound-Gaussian clutter modelling with an inverse Gaussian texture distribution," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 876–879, Dec. 2012.

[12] E. Ollila and E. Raninen, "Optimal shrinkage covariance matrix estimation under random sampling from elliptical distributions," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2707–2719, May 2019.

[13] J. R. Magnus and H. Neudecker, "The commutation matrix: Some properties and applications," *Ann. Statist.*, vol. 7, no. 2, pp. 381–394, 1979.

[14] E. Ollila, D. P. Palomar, and F. Pascal, "Shrinking the eigenvalues of M-estimators of covariance matrix," *IEEE Trans. Signal Process.*, vol. 69, pp. 256–269, 2021.

[15] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.

[16] D. E. Tyler, "Radial estimates and the test for sphericity," *Biometrika*, vol. 69, no. 2, pp. 429–436, Aug. 1982.

[17] J.-F. Cardoso, "Source separation using higher order moments," in *Proc. Int. Conf. Acoust., Speech Signal Process.*, 1989, pp. 2109–2112.

[18] D. E. Tyler, F. Critchley, L. Dümbgen, and H. Oja, "Invariant co-ordinate selection," *J. Roy. Stat. Soc.: Ser. B (Stat. Methodol.)*, vol. 71, no. 3, pp. 549–592, 2009.

[19] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, 2004.

[20] O. Ledoit and M. Wolf, "Honey, I shrunk the sample covariance matrix," *J. Portfolio Manage.*, vol. 30, no. 4, pp. 110–119, Jul. 2004.

[21] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 46, no. 1, pp. 449–458, Jan. 2010.

[22] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero, "Shrinkage algorithms for MMSE covariance estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 10, pp. 5016–5029, Oct. 2010.

[23] J. Bun, R. Allez, J.-P. Bouchaud, and M. Potters, "Rotational invariant estimator for general noisy matrices," *IEEE Trans. Inf. Theory*, vol. 62, no. 12, pp. 7475–7490, Dec. 2016.

[24] O. Ledoit and M. Wolf, "Nonlinear shrinkage estimation of large-dimensional covariance matrices," *Ann. Statist.*, vol. 40, no. 2, pp. 1024–1060, Apr. 2012.

[25] D. T. Searls and P. Intarapanich, "A note on an estimator for the variance that utilizes the Kurtosis," *Amer. Statistician*, vol. 44, no. 4, pp. 295–296, 1990.

[26] L. A. Goodman, "A simple method for improving some estimators," *Ann. Math. Statist.*, vol. 24, no. 1, pp. 114–117, Mar. 1953.